Journal of Cleaner Production 174 (2018) 1629–1641

Contents lists available at ScienceDirect

Journal of Cleaner Production

journal homepage: www.elsevier.com/locate/jclepro

Assessing urban areas vulnerability to pluvial flooding using GIS applications and Bayesian Belief Network model

Yekenalem Abebe [a, *], Golam Kabir [b], Solomon Tesfamariam [a]

[a] School of Engineering, University of British Columbia (UBC), 3333 University Way, Kelowna, BC, V1V1V7, Canada
[b] Department of Mechanical, Automotive, Materials Engineering, University of Windsor, CEI-2162, 401 Sunset Avenue Windsor, ON, N9B 3P4, Canada

ARTICLE INFO

Article history:
Received 29 April 2017
Received in revised form
18 October 2017
Accepted 10 November 2017
Available online 13 November 2017

Keywords:
Pluvial flood
Vulnerability
Urban drainage
Basement flooding
Geographic Information System (GIS)
Bayesian Belief Network (BBN)

ABSTRACT

Expected increases in intensity and frequency of rainfall extremes due to climate change, and increased paving and loss of water storage space in urban areas is making cities more susceptible to pluvial flooding. Evaluating the flood vulnerability of urban areas is a crucial step towards risk mitigation and adaptation planning. In this study, a coupled Geographic Information System and Bayesian Belief Network based flood vulnerability assessment model is proposed. The methodology can quantify uncertainty and capture the casual nexus between pluvial flood influencing factors. The model is applied in a case study to diagnose the reason behind the variations in the number of reported basement flooding in different parts of the City of Toronto and to predict Flood Vulnerability Index (FVI). The predicted FVI is validated by comparing the results with the number of approved basement flood subsidy protection program applications. The case study result shows that areas located near downtown Toronto have high FVI and most of the city has medium to low FVI.

1. Introduction

Prevalent urban flooding in urban areas is increasing because of growing population with inappropriate land use planning and climate change-induced rise in extreme rainfall events (Karamouz et al., 2014; Liao, 2012; Saul et al., 2010). Flooding affects a wide range of stakeholders in urban areas. Restoring flooding and associated sewer backup damages is expensive for property owners, municipalities, and insurance companies. A report by Swiss Re estimates that severe flood events in Canada can cause losses exceeding $13 billion and less than half of these is covered by insurance (Swiss Re, 2013). Therefore, it is crucial to advance flood risk management strategies to reduce the impacts of future flood events.

Urban flood types can be categorized into pluvial, groundwater, fluvial (also called as riverine), and coastal flooding (Hammond et al., 2015; Saul et al., 2010). Pluvial flooding occurs due to inefficient and inadequate urban drainage system with respect to rainfall intensity. Other types of floods reported include flash floods, sewer derived, tsunami, dam break and canal breach (Foody et al., 2004; Kim and Sanders, 2016; Van Veen et al., 2014).

Flood risk is the product of a hazard and its consequences (Kron, 2005). Flood hazard can be quantified in terms of flood depth, duration, velocity, sediment contains and its probability. While the consequence depends on the exposure (number of people and infrastructure impacted) and the vulnerability and resiliency of the affected area (Fig. 1). Holling (1973) first presented the idea of resilience as the ability of a system to deal with severe changes and remain to function. Vulnerability can be understood as the propensity or predisposition of a system to be adversely affected (Lavell et al., 2012). Therefore, urban areas flood risk depends on the systems components vulnerability and resiliency to flooding and the likelihood and intensity of flood hazard in the area (Fig. 1).

Flood vulnerability assessment approaches differ in their scope, theoretical framework, variables considered and analysis scale. Nasiri et al. (2016) group the methodologies into vulnerability indicators methods, curve method, disaster loss data method, and modeling methods. Balica et al. (2013) clustered them into physically based and parametric models. A majority of the proposed methodologies deal with the physical aspect of vulnerability (Blanco-Vogt and Schanze, 2014; Pradhan, 2010; Rahmati et al., 2016; Tehrany et al., 2015b) and some provide a holistic approach that considers the economic and social dimensions (Balica et al., 2012, 2009; Kubal et al., 2009; Nitivattananon et al., 2009).

* Corresponding author.
E-mail addresses: yekenalem.abebe@ubc.ca (Y. Abebe), golam.kabir@uwindsor.ca (G. Kabir), solomon.tesfamariam@ubc.ca (S. Tesfamariam).

https://doi.org/10.1016/j.jclepro.2017.11.066
0959-6526/

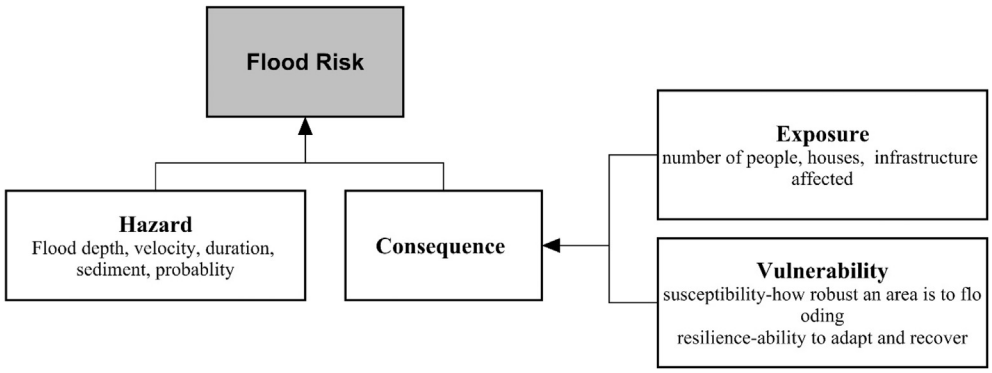


Fig. 1. Flood risk components.

Furthermore, most papers are focused on fluvial and coastal flooding and few papers discuss pluvial flood vulnerability. Moreover, analysis scale and detail vary widely. In general, the available vulnerability assessment methodologies can be grouped into physically based, data driven and expert knowledge based models.

Physically based pluvial inundation modeling techniques are evolving towards coupled 1D-2D simulation models (Van Dijk et al., 2014). Data-driven models consist of statistical methods, machine learning, and ensemble learning methods. These are empirical models using parameters that directly or indirect affect flood occurrence to estimate the flood susceptibility of an area (Pradhan, 2010; Tehrany et al., 2014a, 2013). Vulnerability curve methods (empirical damage or fragility curves) provide the correlation between the asset at risk and flood hazard.

Studies that proposed an indexing system for flood vulnerability and resilience assessment are usually based on experts' knowledge. The majority of these models considered only weighted sum (Batica and Gourbesville, 2014; Karamouz et al., 2014) or weighted mean index (Cutter et al., 2010; Shaw, 2009) aggregation methods. Both methods assume that the parameters considered to be independent. Orencio and Fujii (2013) implemented analytic hierarchy process and weighted linear average to prepare a disaster-resilience index. Similar methodologies were used by Stefanidis and Stathis (2013) and Zou et al. (2013) to prepare flood hazard and flood vulnerability indexes. Recently, Sun et al. (2016) proposed a flood resilience assessment method using analytic network process.

It is recognised that urban flood vulnerability and resilience factors are interdependent and there exists non-linear relationship. The identification and consideration of the interrelationship between flood causative factors are vital for vulnerability and resiliency assessment. The empirical and knowledge based models discussed above fail to consider these interdependencies and do not consider the uncertainty in the analysis, which is a vital element for decision-making. This study presents a methodology to compute FVI that can quantify uncertainty and capture the casual nexus between pluvial flood influencing factors. The proposed methodology is discussed in detail in the next section. Section three and four present the case study data collection and results respectively. The conclusion is summarized in section five.

2. Methodology

To improve the main shortcomings of currently available flood vulnerability assessment approaches, a coupled GIS and BBN-based model that can capture the casual nexus between factors, capable of quantifying uncertainty and that can utilize both data and knowledge-based sources is proposed. The proposed methodology workflow is shown in Fig. 2.

2.1. Bayesian Belief Network (BBN)

Bayesian belief network model is supported by a graphical network representing cause and effect relationships between different factors considered in a study (Pearl, 1988). The qualitative component of a BBN is a directed acyclic graph, where nodes and directed links signify system variables and their causal dependencies (Cockburn and Tesfamariam, 2012; Jensen and Nielsen, 2007; Pearl, 1988). Fig. 3 shows a sample BBN model where node *X* and node *Y* are the parents of the child node *Z* for the directed links from *X* and *Y* to *Z*.

The quantitative component of a BBN model is presented with a set of conditional probabilities or probability distributions for each child node given its parent nodes in the network (Pearl, 1988; Thomsen et al., 2016). As BBN model generally deals with discrete probabilities, each node is classified into a finite set of state values accompanying with a probability (Jensen and Nielsen, 2007). The dependencies between child node *Z* with parent nodes *X* and *Y* are quantified by a conditional probability table (CPT), as demonstrated in Fig. 3. For the nodes with no parents like *X* and *Y* in Fig. 3, the CPT transfers to the unconditional probability (UP). The CPTs and UPs can be attained through expert decision or knowledge elicitation (Joseph et al., 2010; Nadkarni and Shenoy, 2001) and/or training from data (Kabir et al., 2015; Tang and McCabe, 2007).

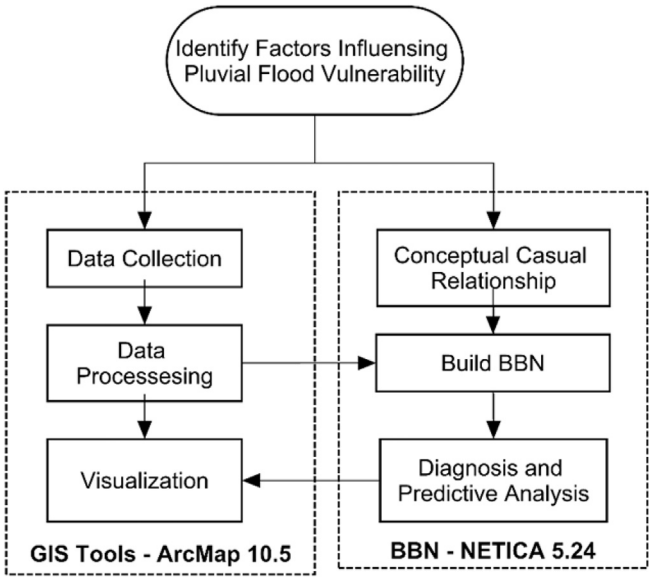
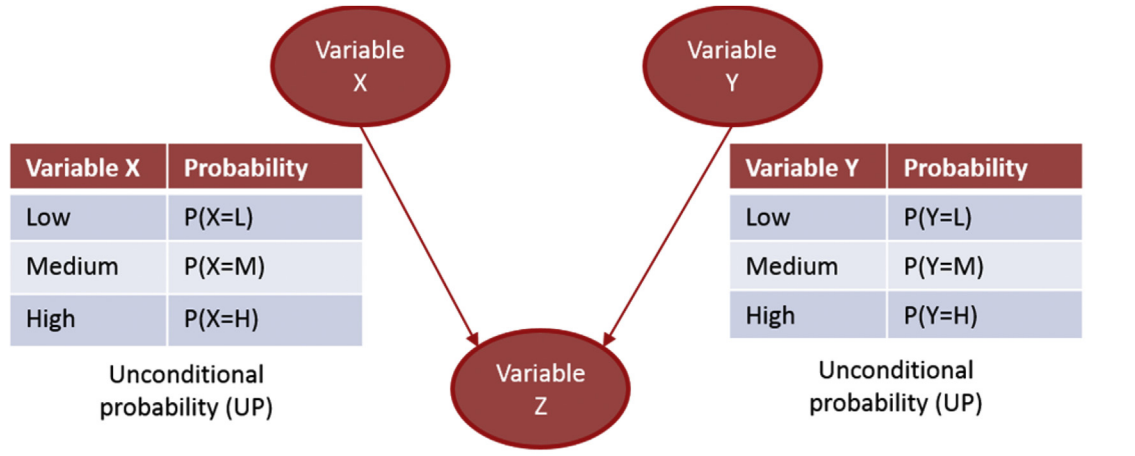


Fig. 2. Proposed methodology workflow.



Variable X	Probability
Low	P(X=L)
Medium	P(X=M)
High	P(X=H)

Unconditional probability (UP)

Variable Y	Probability
Low	P(Y=L)
Medium	P(Y=M)
High	P(Y=H)

Unconditional probability (UP)

Variable A_1	Variable A_2	Variable Z		
		Probability		
		L	M	H
L	L	P(Z=L|X=L, Y=L)	P(Z=M|X=L, Y=L)	P(Z=H|X=L, Y=L)
...	...	...	...	...
H	M	P(Z=L|X=H, Y=M)	P(Z=M|X=H, Y=M)	P(Z=H|X=H, Y=M)
H	H	P(Z=L|X=H, Y=H)	P(Z=M|X=H, Y=H)	P(Z=H|X=H, Y=H)

Conditional probability table (CPT)

Fig. 3. Representation of a BBN.

One of the main advantages of BBNs is that it performs probabilistic inferences or belief updating based on data or observations using Bayes' theorem (Jensen and Nielsen, 2007). In a BBN analysis, the updated probability for the m number of mutually exclusive variables or parameters $V_{i\ (i = 1,2, \ldots,m)}$, and given evidence or data D, is determined by the following relationship:

$$p(V_j|D) = \frac{p(D|V_j) \times p(V_j)}{\sum_{i=1}^{m} p(D|V_i) \times p(V_i)} \quad (1)$$

where $p\ (V_j|D)$ is the posterior probability of V based on the data or evidence D; $p\ (V_j)$ refers the prior probability; $p\ (D|V_j)$ denotes the conditional probability considering V_j is true and the denominator is a constant value represents the total probability (Pearl, 1988). BBNs are flexible enough to perform bottom-up inference or diagnostic analysis (from Z to X, i.e., reasoning the possible cause observing the effect) and top-down inference or predictive analysis (from X to Z, i.e., reasoning the possible effect observing the cause) (Cockburn and Tesfamariam, 2012; Jensen and Nielsen, 2007; Thomsen et al., 2016). A BBN model can also provide informed decisions efficiently in case of imprecise, incomplete or ambiguous data or information (Cooper and Herskovits, 1992).

2.2. BBN-based flood vulnerability model

To construct the BBN model, important factors that can contribute to the occurrence of urban floods are identified. These factors are presented as parent nodes in the BBN model (Fig. 4). The parameters used are land cover, population density, slope, soil drainage class, drainage density, DEM, rainfall, and drainage capacity. Urbanization has profoundly altered our citys' land cover. Urban catchments have very low permeability and high rainfall – runoff conversion rates. These changes affect the hydrology that determines flood hazard (Fratini et al., 2011, 2012). Drainage infrastructure capacity, type and condition are other key factors influencing the occurrence of pluvial floods. Urban drainage infrastructures are composed of two sub-systems: minor and major systems. The minor system refers to an underground sewer network while the overland flow route is called the major system (Maksimović et al., 2009). These infrastructures have to be in good condition and perform according to their design requirements to prevent pluvial flooding. Therefore, not only design capacity have to be considered in flood vulnerability assessment, but also the current condition and performance.

Design guidelines, that dictate the required design capacity of drainage systems, are other important factors that affects the likelihood of pluvial flood occurrence. Such standards are provided by local municipalities and the critical design rainfall events depends on the required level of protection (return period). For example, Toronto's municipal design guideline recommends return period ranging from 2 years up to 100 years based on the type of road (local, collector, arterial, highway, road underpass) and drainage system (minor, major) (Klimas, 2009). Furthermore, micro-topography effects can cause localized drainage system failures (Aronica and Lanza, 2005). Masoudian and Theobald (2011) showed the effect of topography on flood parameters, such as maximum flood discharge and time to peak. Valeol and Rasmussen (2000) used topographic index distribution to understand the

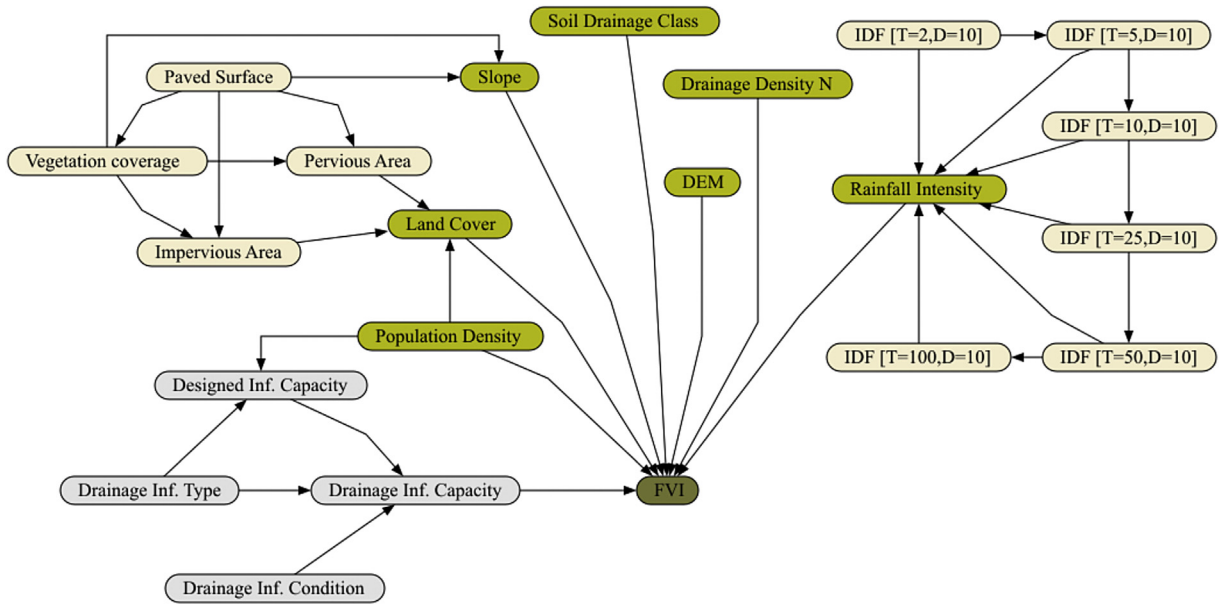


Fig. 4. BBN-based pluvial flood vulnerability model.

Table 1
Parameters used in previous studies.

No	Articles	Land cover	Slope	DEM	Soil Infiltration	Precipitation	Drainage system	Natural drainage	Population density
1	(Pradhan, 2010)	✓	✓	✓	✓	✓			
2	(Wang et al., 2011)	✓		✓		✓	✓		✓
3	(Chen et al., 2011)			✓		✓	✓		
4	(Lee et al., 2012)	✓	✓	✓	✓			✓	
5	(Tehrany et al., 2014b, 2013)	✓	✓	✓	✓	✓		✓	
6	(Tehrany et al., 2015a)		✓	✓	✓			✓	
7	(Tehrany et al., 2015b)	✓	✓		✓	✓		✓	
8	(Rahmati et al., 2016)	✓	✓	✓	✓		✓		

effect of topography on a small catchment's flood frequency distribution in Ontario. These finding showed that flood parameters are sensitive to topography changes and topographic distributions have a considerable influence on peak runoff rates. Previous studies that used a combination of these parameters in flood vulnerability assessment are summarized in Table 1.

The BBN model shown in Fig. 4 is constructed using the commercially available Netica software package. Tree Augmented Naive Bayes (TAN) structure learning is used to validate the link structure based on data collected from Toronto. Netica has three algorithms (i.e., counting algorithm, expectation-maximization (EM) and gradient descent) to learn the conditional probabilities in the network (Neapolitan, 2004). In our case study, there is an incomplete data for the soil infiltration capacity and we have two hidden variables (land cover and rainfall intensity). Therefore, the EM and gradient descent algorithms were used and the results are discussed in section 4. The factors classification is summarized in the Tables 2–5 below.

Table 2
Land use factors classification.

Factor	Class	Factor	Class
Paved Surface [%]	0–0.19	Pervious area [%]	0.1–0.42
	0.19–0.36		0.42–0.56
	0.36–0.47		0.56–0.67
	0.47–0.62		0.67–0.82
	0.62–1.00		0.82–1.00
Vegetation coverage [%]	0–0.35	Land cover	0–0.48
	0.35–0.5		0.48–1.01
	0.5–0.62		1.01–1.94
	0.62–0.79		1.94–4.05
	0.79–1.00		4.05–9
Impervious area [%]	0–0.17	Population density [per Km^2]	1706–2619
	0.17–0.32		2619–3622
	0.32–0.42		3622–4720
	0.42–0.56		4720–6494
	0.56–0.9		6494–9487

Table 3
Hydrological factors classification.

Factor	Class	Factor	Class
Soil drainage	Well drained	Drainage Density [%]	0–0.0067
	Imperfectly drained		0.0067–0.0177
	Poorly drained		0.0177–0.028
	Very poorly drained		0.028–0.041
	Variable		0.041–0.074
DEM [M]	71–100	Rainfall Intensity [mm]	12.57–13.87
	100–128		13.87–14.64
	128–152		14.64–15.42
	152–173		15.42–16.51
	173–210		16.51–18.74

Table 4
IDF curves classification.

Factor	Class	Factor	Class
IDF [T = 2, D = 10] [mm]	10.71–11.91	IDF [T = 25, D = 10] [mm]	17.06–21.89
	11.91–12.04		21.89–22.50
	12.04–12.23		22.50–22.81
	12.23–12.58		22.81–24.45
	12.58–15.09		24.45–29.76
IDF [T = 5, D = 10] [mm]	13.43–15.99	IDF [T = 50, D = 10] [mm]	18.97–26.88
	15.99–16.25		26.88–27.68
	16.25–16.46		27.68–28.14
	16.46–17.38		28.14–30.37
	17.38–20.94		30.37–37.00
IDF [T = 10, D = 10] [mm]	15.15–18.61	IDF [T = 100, D = 10] [mm]	20.58–26.88
	18.61–19.00		26.88–27.68
	19.00–19.24		27.68–28.14
	19.24–20.52		28.14–30.37
	20.52–24.71		30.37–37.00

Table 5
FVI classification.

Factor	Scale		Class (Based on RBF)	Class (Based Approved BFSPP)
FVI	Very low	1	<19	<17
	Low	2	19–36	17–33
	Medium	3	36–55	33–49
	High	4	55–80	49–65
	Very high	5	80–136	65–82

The FVI of a grid cell is calculated based on the number of reported basement flooding (RBF). While the approved basement flood subsidy protection program (BFSPP) is used to validate the FVI prediction output according to the classification in Table 5.

2.3. GIS tools

GIS software, ArcGIS 10.5 tools are used for data collection, processing, and result visualization. The study area is classified into small grid cells and each grid cell is represented by a single row in the maps attribute table. The attribute table raw size is the same as the number of grid cells and the columns are equal to the number of parameters considered in the BBN. Each grid cell has its own unique data that gives evidence regarding the causal relationships between the parameters considered and number of RBF. We have classified The City of Toronto into 760 grid cells. Except for zones near the city's boundary, all grid cells have an area of one square kilometer. The grid cell size must be determined based on available raw data resolution and required level of detail in the analysis. Smaller grid cell maximizes the number of training case files (evidence) for the BBN. Therefore, it is preferable to use the highest possible resolution to get a more localized result.

Raw data for the model input parameters could be collected from different sources in the form of maps, tables, and publications. Some of the maps are provided in raster or shape files format that can directly be imported into GIS software platform, however, other require additional conversion process and georeferencing. Once the raw data is converted into the required format, it has to be processed and summarized into the grid cells. This processes can be classified into two: upscaling and downscaling. Downscaling is required when the raw data's resolution is lower than the analysis grid cells (>1 km^2). Conversely upscaling is needed if the resolution of the data is higher (<1 km^2). It is important to understand that this process generalizes or extrapolate the raw data into each grid cells and the processed data quality could be affected. Therefore, the data processing step has to be supported by theoretical reasoning and the consistency of the results should be verified by experts' knowledge. Furthermore, the grid cells size should be defined in order to minimize data processing requirements.

3. Case study

In this section, the methodology proposed is applied to investigate the pluvial flood vulnerability of the City of Toronto, as defined in Fig. 5a. Toronto is the fourth most populous city in North America and the largest city in Canada with about 2.8 million inhabitants (City of Toronto, 2016a). The population in the Greater Toronto Area (which includes Durham, Halton, Peel and York) is expected to increasing by over 42.9 percent over the next three decades and its drainage infrastructure needs to cope up with the continuous service demand growth (City of Toronto, 2016a). Additionally, as one of the oldest city in Canada, aging-related infrastructure deterioration poses a major challenge to the sewer systems reliability. Currently, the city has a $1.6 billion state of good repair backlog (Kellershohn, 2016). The data collection, processing, and analysis steps for the model development process is discussed in the following sections.

3.1. Data collection

Twenty-one parameters related to topographic, climatologic, soil, drainage infrastructure, watershed and land cover are used to construct the BBN (Fig. 4). The data collection and processing steps for these parameters are discussed below.

3.1.1. Administration and population density

The City of Toronto covers a total area of around 641 sq.km and it is divided into planning districts and wards (City of Toronto, 2016a). It has four districts with three planning sections each, except for Toronto & East York district, which has four sections (refer Fig. 5a). Furthermore, Toronto is divided into 44 wards with an average area of 15 km^2 and population of about 61,000 (City of Toronto, 2014). Fig. 5b shows the population density for the different wards in the city. Wards located near downtown Toronto have the highest population density with more than 9000 people per square kilometer while areas near the city periphery have pollution density below 3000.

3.1.2. Drainage infrastructure and basement flooding

Toronto's urban drainage infrastructure consists of two separate systems: the minor system and the major system (Klimas, 2009). The minor system is designed to convey low intensity rainstorms that are more frequent. Most drainage systems in older neighborhoods only have the minor system, that increases the likelihood of basement flooding in those areas (Klimas, 2009). The City of Toronto – open data catalog contains basement flood events reported by districts and wards from January 2013 to June 2015 (City of Toronto, 2016b). According to the data, in 2013, there were more

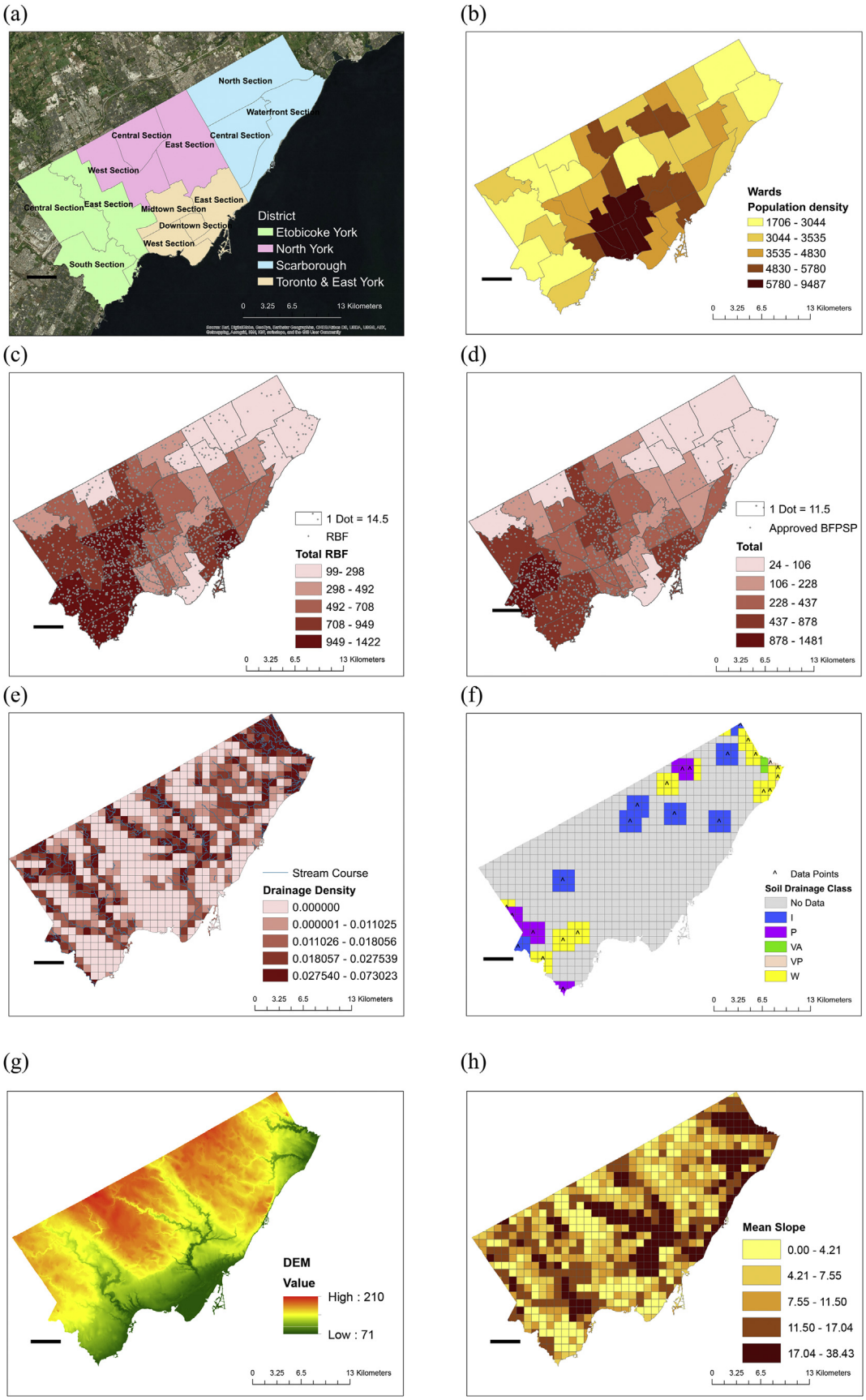


Fig. 5. Data collection and processing: (a) city boundary (b) population density (c) RBF (d) BFSPP approved (e) River and stream network (f) Soil drainage class (g) DEM (h) Mean slope.

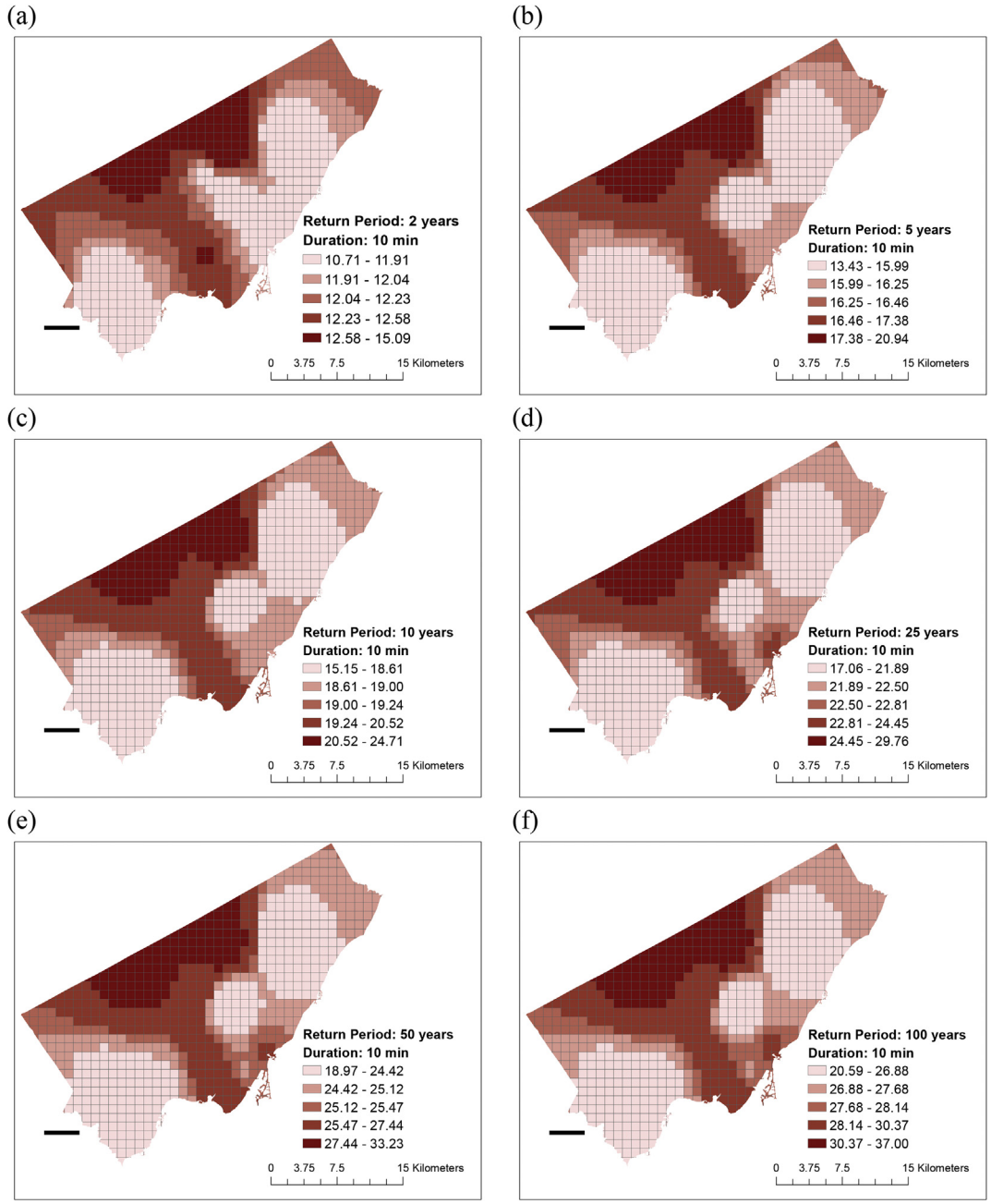


Fig. 6. Data collection and processing: Average IDF for (a) T = 2, D = 5 min (b) T = 5, D = 10 min (c) T = 10, D = 10 min (d) T = 25, D = 10 min (e) T = 50, D = 10 min (f) T = 100, D = 10 min.

than fourteen thousand RBF and around five thousand applications were approved for BFSPP. The maps in Fig. 5c and d shows the sum of all RBF and approved BFSPP applications, respectively.

Wards located in the western part of the city (Etobicoke York district) reported the highest number of basement flooding and eastern part of the city (Scarborough) registered the lowest. To reduce basement flooding, the City has an ongoing basement flooding protection program to construct recommended infrastructure improvements in flood prone areas. Furthermore, the BFSPP provides up to $3400 support for homeowners to make important improvements to reduce the risk of basement flooding by install flood protection devices (City of Toronto, 2016c). Wards located in Etobicoke York district received the highest number of approval for BFSPP while Scarborough district received the lowest.

3.1.3. *Watershed*

Toronto lies under the larger Great Lakes Basin and it has six river watersheds and the waterfront catchment that drains into Lake Ontario (EverGreen, 2017). Since natural stream courses can be considered part of the cities major drainage system, based on the stream course dataset available from the city's open data catalog, the total length of streams crossing each study grid cells is calculated ("natural drainage density") and summarized in Fig. 5e.

3.1.4. *Soil*

Soil drainage class data from Land Information Ontario (LIO) was used to extract soil drainage class data. According to the National Soil DataBase (NSDB), there are seven soil classes ranging from very

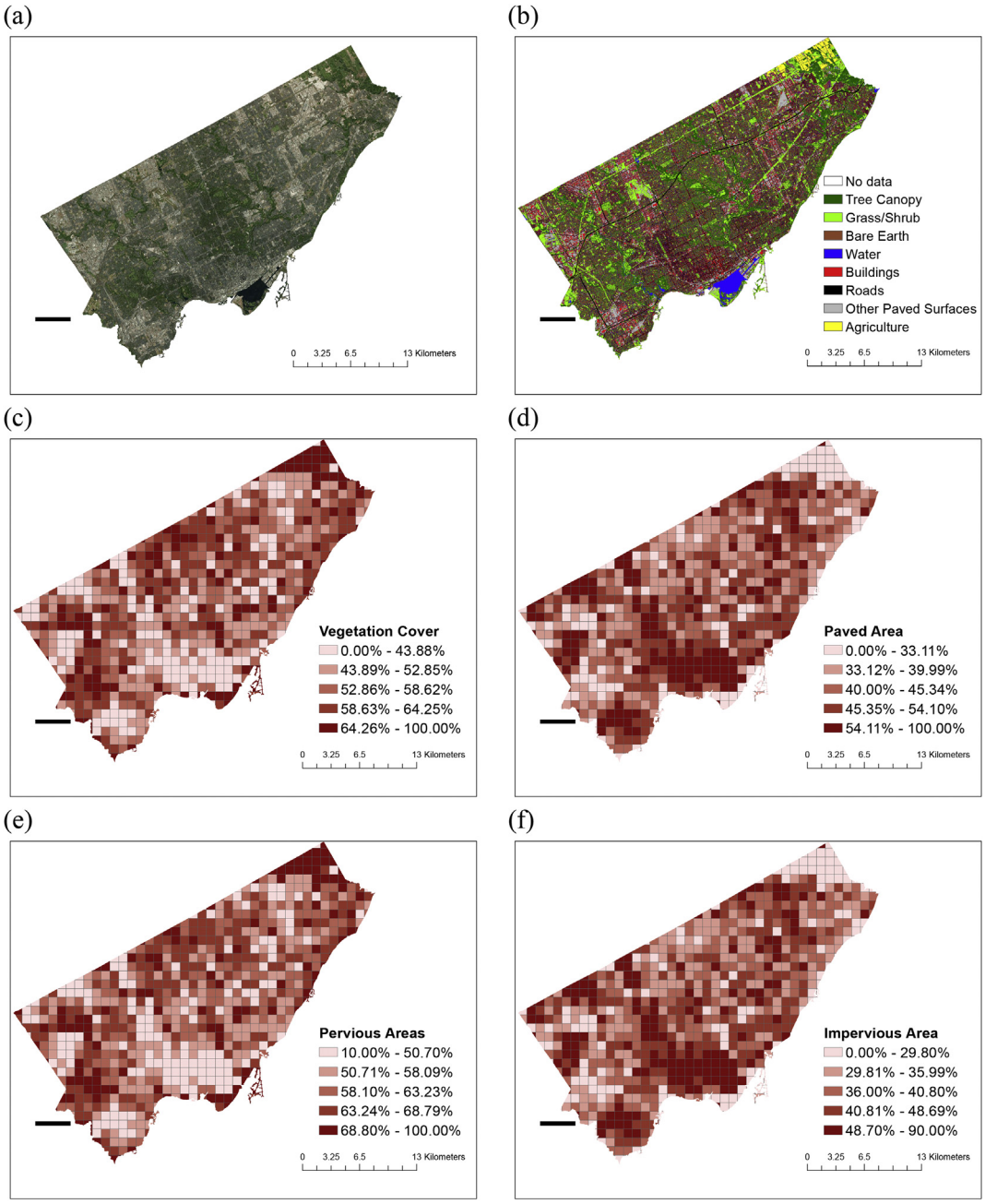


Fig. 7. Data collection and processing: Land cover (a) google earth image (b) land cover classification: source (c) Average vegetation cover (d) Average paved area (e) Average pervious area (f) Average impervious area.

rapidly drained to very poorly drained soils. Since soil type data is available only for some parts of the city, areas in near proximity (1 km radius) from available data points are assumed to have the same drainage class (refer to Fig. 5f).

3.1.5. Topography

The Canadian Digital Elevation Model (CDEM) prepared by Natural Resources Canada is used to calculate the average slope in each study grid cells. Toronto is mostly flat with altitude ranging from 71 m gently increasing upward away from the shore up to 210 m. Sixty percent of the average slopes are below 11.5 percent and only twenty percent of the areas have slopes higher than 17 percent (Figs. 5g and 4h).

3.1.6. Climate

We have considered a total of six Intensity-Duration-Frequency curves (IDF), with 2, 5, 10, 25, 50 and 100 years return period and rainfall duration of 10 min. Engineering Climate Datasets, published by Environment and natural resources (2017), is used to calculate the IDF values for the study area. There are 12 IDF stations in and around the city of Toronto. Based on these stations, the IDF values for all part of the city is calculated using Inverse Distance Weighted (IDW) approach. Finally, the average IDF value for each grid cells is calculated and presented in Fig. 6.

3.1.7. Land cover

Data obtained from the city's open data catalog is used to analyze the forest and land cover of the study area (City of Toronto,

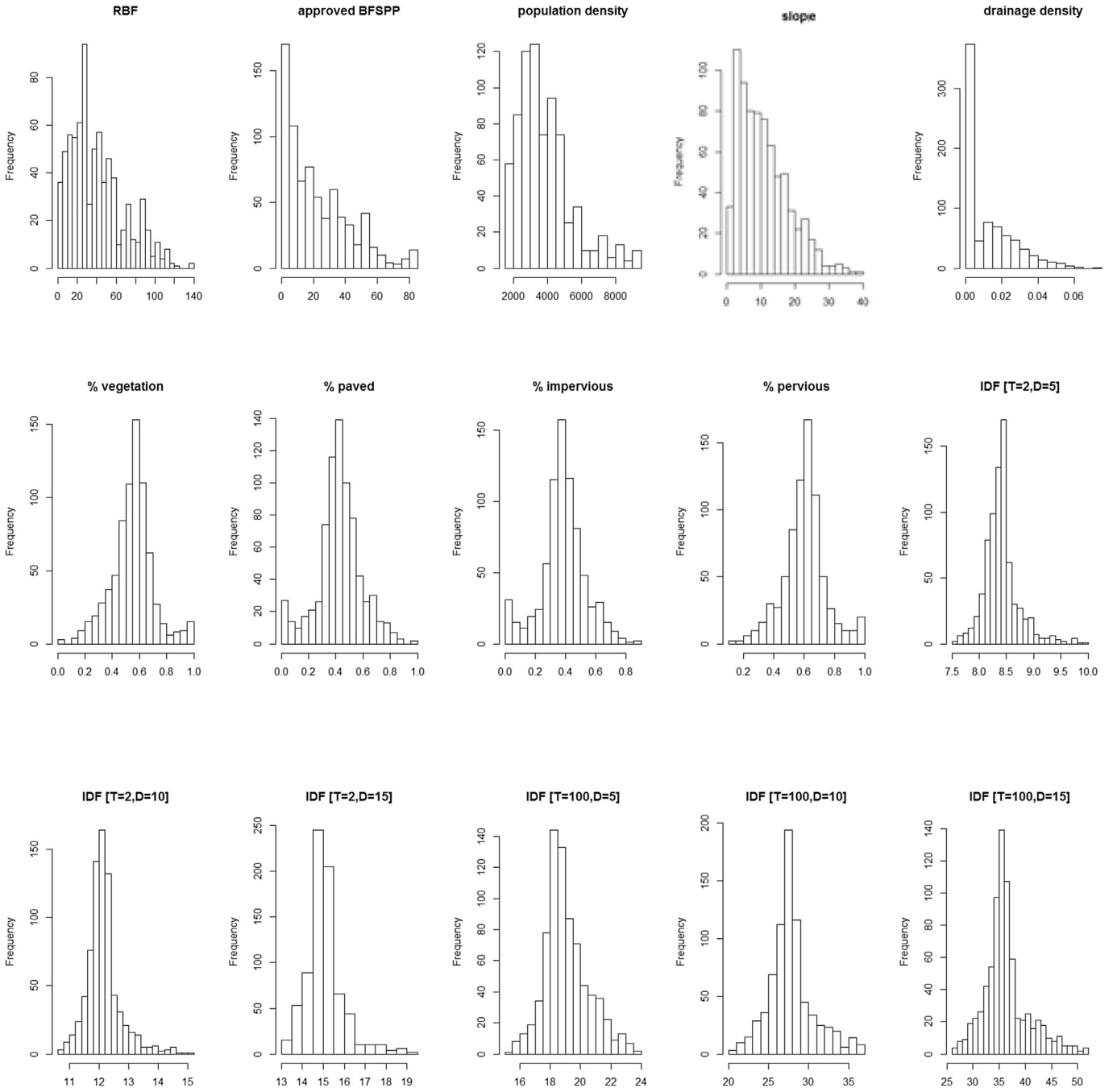


Fig. 8. Histogram of inputs parameters - the frequency distribution of the 760 cells in the analysis grid.

2016b). Of particular interest for our case study are the relative proportion of green areas and paved surfaces. The dataset has eight land cover classes as shown in Fig. 7b. From these, we have estimated the total area of green area, paved surface, pervious areas and impervious areas for each grid cell in the study area. Green area (vegetation cover) is calculated as the sum areas for tree canopy, grass/shrub, and agriculture areas, whereas paved surfaces constitute buildings, roads, and other paved surfaces. Pervious areas is estimated by summing up areas for green area and adding 10 percent of the paved surface area. The remaining 90 percent of the paved surface is assumed to be the impervious area and finally the relative percentage of each parameter is summarized in Fig. 7c–f.

The resulting maps may not represent the exact ground condition because of two crucial shortcomings. Firstly, the raw dataset used represents a "top down" mapping approach where tree canopy overhanging other structures is grouped into the tree canopy class (Urban Forestry, 2009). This can significantly increase the vegetation cover area. Secondly, missing data and gaps in the dataset could distort the final values in some part of the city. However, the analysis gives us a vital comparative information regarding the parameters, which affects infiltration and runoff generation rate, across the study area. Google imagery, raw land cover data, and the resulting four parameters map are displayed in Fig. 7 below. Overall, it can be seen that downtown areas have low vegetation cover, low pervious areas, and high paved surface area.

The histogram for the parameters considered in the model and

their classification is plotted in Fig. 8 and Table 2, respectively. The graphs for the number of approved BFSPP applications, number of RBF, average slope, average population density and drainage density shows a right skewed distribution that implies most of the grid cells in the study area have values lower than the mean. This trend is consistent with the characteristics of the study area. For example, only areas near downtown Toronto have a high population density that explains the right skewed data for average population density. The slope histogram shows similar trend because the study area is generally flat and only a few areas have a slope greater than 17 percent. Even though the data for the remaining land use and rainfall intensity related parameters are asymmetric, they resemble normal distribution with most of the data points concentrated near the mean (see Table 3).

4. Results and discussion

4.1. Sensitivity analysis

Sensitivity to finding report for a node provides the connected nodes (directly or indirectly connected) level of influence on its value. The sensitivity of a node can be quantified by variance reduction, mutual information, or variance of beliefs (Norsys Software Corp, 2016). Variance reduction is best suited for a continuous nodes while mutual information (i.e. entropy reduction) can be used for discreet nodes. The variance and expected value of a continuous (or discrete) node Q given the evidence f can be computed as:

$$V(Q/f) = \sum_{n=1}^{q} P(q/f)\left[X_q - E(Q/f)\right]^2 \quad (2)$$

$$E(Q/f) = \sum_{n=1}^{q} P(q/f)X_q \quad (3)$$

where q is the state of the query node Q, f is the state of the varying node F; Xq is the state number (numeric real value) corresponding to state q; $E(Q)$ is the expected real value of Q before any new findings; $E(Q/f)$ is the expected real value of Q after a new finding f for node F; $P(q)$ and $P(q/f)$ are the probabilities of state q before and after a finding, respectively (Norsys Software Corp, 2016).

Table 6 presents the sensitivity to finding report for the node FVI. The list summarizes the parameters according to their order of influence and their level of influence. For this study, we are most interested in the order of influence and the result reveals four important points.

- Population density is the most influential factor affecting FVI. Densely populated areas near downtown Toronto recorded a higher number of basement flooding (Fig. 5). Population density has direct influence on land cover and other urban areas density indicators such as floor area ratio, residential density and building density. For high-resolution analysis, these additional nodes can be incorporated in the model.
- Land cover related parameters and slope are the second groups of most influential factors affecting the number of RBF. Areas with a high proportion of vegetation cover and pervious area reported a lower number of basement flooding while built up areas near downtown Toronto recorded a higher number of basement flooding.
- Although there is a variation in rainfall intensity in different parts of the city, the impact it has on the uneven number of reported basement flooding is minimal. For example, areas with relatively higher IDF values in the Northern York district have lower RBF, this could be due to their higher vegetation cover and lower population density. However, this does not mean that rainfall intensity is not an important factor in pluvial flooding process. Our result only shows that the climate variations within the city is not the main reason driving the uneven number of reported basement flooding.

Table 6
Sensitivity of the FVI node to a finding at another node.

Node	Variance Reduction
Population Density	13.63
Land Cover	2.273
Slope	1.674
DEM	1.621
Impervious Area	1.528
Paved Surface	1.489
Pervious Area	1.351
Drainage Density	1.293
Vegetation cover	1.019
Rainfall Intensity	0.9838
IDF [T = 100 D = 10]	0.1123
IDF [T = 50 D = 10]	0.1048
IDF [T = 25 D = 10]	0.1021
IDF [T = 2 D = 10]	0.09964
IDF [T = 5 D = 10]	0.09862
IDF [T = 10 D = 10]	0.08455
Soil Drainage Class	0.08071

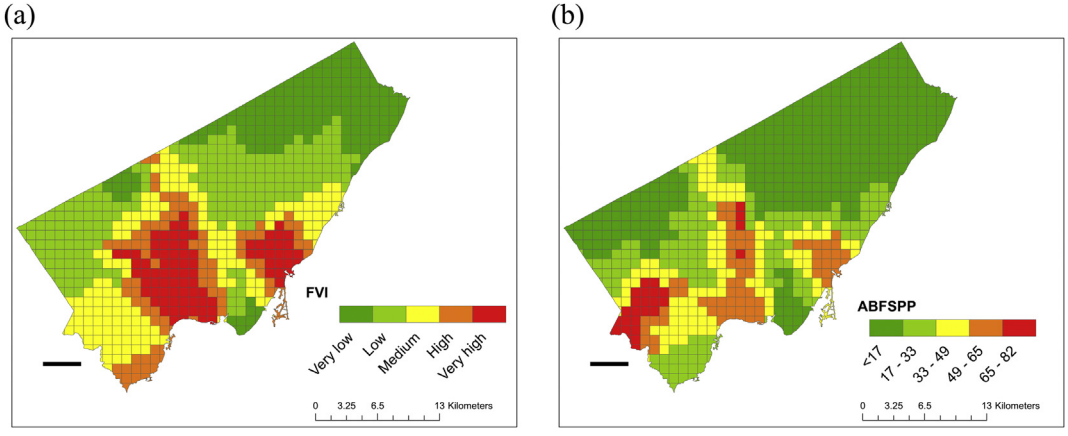


Fig. 9. (a) Expected FVI and (b) Approved BFSPP.

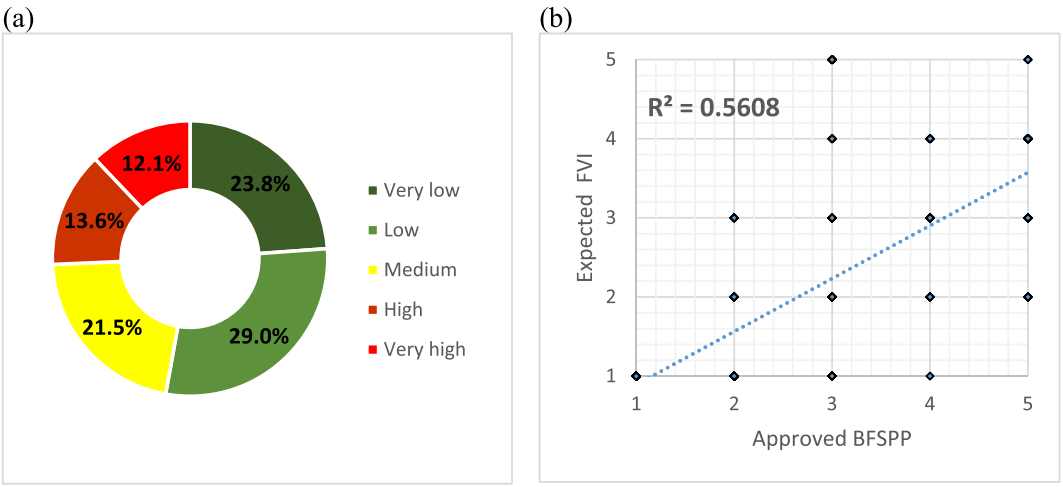


Fig. 10. (a) Percent distribution of FVI result for the city of Toronto (b) Expected FVI vs. Approved BFSPP.

- In terms of order of influence for the FVI node (RBF), longer return period storm events are more influential than shorter return period storm event.

4.2. Flood vulnerability index

After diagnosing the reason behind the disparity in the number RBF, using the developed model, the expected value of FVI for all the 760 study grid cells is predicted and summarized in Fig. 9a, whereas the average number of approved BFSPP in each grid cell is shown in Fig. 9b.

More than 70 percent of the grid cells have medium or lower FVI and only a quarter of the grid cells record a very high or high FVI to pluvial flooding (Fig. 10a). The predicted FVI is validated by comparing the results with the number of approved BFSPP applications. R squared for the predicted values (expected FVI) and the number of approved BFSPP is 0.56. Better R squared value can be achieved by incorporating missing parameters such as infrastructure condition and using higher resolution input data.

Finally, Fig. 11 shows a sample prediction result for a single study grid cell. The FVI result is given in probability distribution. For instance, grid cell 562 has more than 0.75 probability of having FVI of three or more (high or very high). This makes the proposed model distinct from previous methodologies that quantify FVI deterministically. This gives additional insight for the decision makers into the uncertainty related with the analysis.

4.3. Limitations

The main limitation of the proposed methodology is that its reliability is affected by data availability and data quality. It is

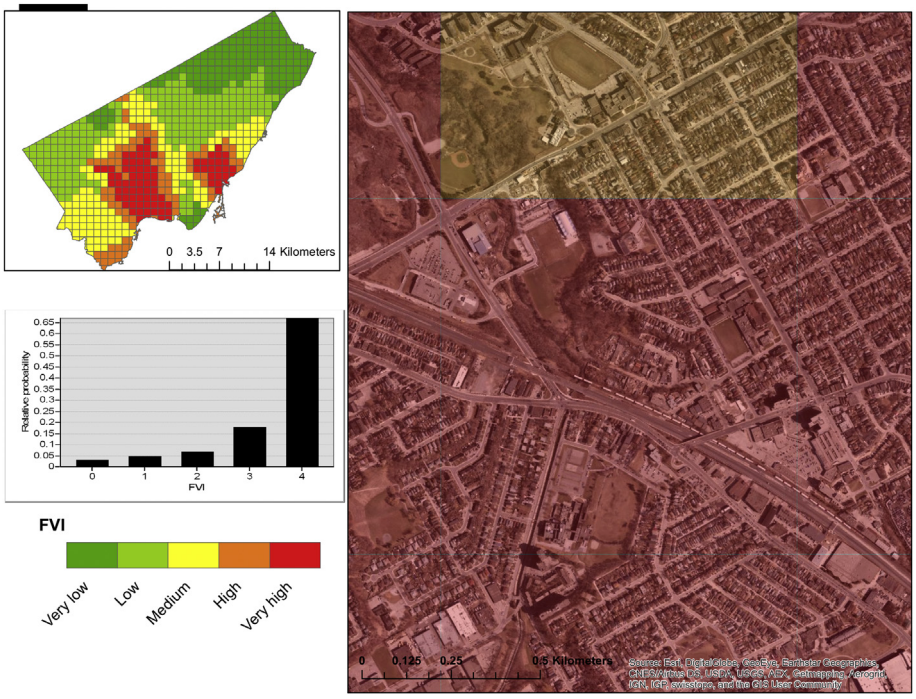


Fig. 11. Example FVI result for a single grid cell [grid code = 562].

important to have a high-resolution data to conduct a more robust investigation. However, in some instances, expert's opinion can be used to substitute missing data. Data availability was a major challenge in the case study discussed. Due to lack of data and low-quality data, drainage infrastructure type, condition and capacity parameters that are shown in BBN were not included in the final analysis. Including these parameters will help us in understanding the full casual nexus between the factors that influence pluvial flooding.

Furthermore, the resolution of the raw data considered and the data processing steps have reduced the quality of the analysis. For example, the RBF data used for the case study was only available by wards; the exact GPS coordinates are unknown. In the future, it is inevitable that advancements in data collection technology will continue to improve data quality, which makes the proposed model very useful. In addition, the BBN model development is only based on one case study. These can undermine the applicability of the model in other municipalities. Therefore, to have a robust model, it is recommended to conduct more case studies and refrain the model accordingly.

5. Conclusions

Understanding the main cause of basement flooding in an area is a crucial step towards flood risk mitigation and adaptation planning. In this study, a novel approach using GIS applications and BBN is proposed to diagnosis the main factors influencing basement flooding and to predict FVI. The case study revealed that climate variations have the least impact on the disproportional number of basement flooding reported by different areas across Toronto. Whereas, population density followed by land cover related parameters are the most influential factors. Therefore, the city of Toronto should focus their resources on land cover related parameters to improve the water retention capacity of the soil and decrease impervious areas. This can be achieved by promoting green development like green roof constructions, encouraging the use of pours pavement materials that allow water infiltration, increasing proportion of green areas in new development projects and so on.

The main advantage of the proposed methodology is that it can be adopted in any urban area, based on its characteristics and data availability and it has the ability to quantify uncertainty in the decision-making process. Unlike previous methodologies where FVI was deterministically quantified, the model proposed estimates probability FVI. Furthermore, the model can perform probabilistic inferences or belief updating when new data or observations become available. Additional factors can be incorporated whenever new data is available and the model can easily be expanded to assess socioeconomic aspects of vulnerability and integrate flood resiliency considerations.

Acknowledgment

The financial support through MITACS and Natural Sciences and Engineering Research Council of Canada (NSERC) Collaborative Research and Development Grant is acknowledged.

References

Aronica, G.T., Lanza, L.G., 2005. Drainage efficiency in urban areas: a case study. Hydrol. Process 19, 1105–1119. https://doi.org/10.1002/hyp.5648.

Balica, S.F., Douben, N., Wright, N.G., 2009. Flood vulnerability indices at varying spatial scales. Water Sci. Technol. 60, 2571–2580. https://doi.org/10.2166/wst.2009.183.

Balica, S.F., Popescu, I., Beevers, L., Wright, N.G., 2013. Parametric and physically based modelling techniques for flood risk and vulnerability assessment: a comparison. Environ. Model. Softw. 41, 84–92. https://doi.org/10.1016/j.envsoft.2012.11.002.

Balica, S.F., Wright, N.G., van der Meulen, F., 2012. A flood vulnerability index for coastal cities and its use in assessing climate change impacts. Nat. Hazards 64, 73–105. https://doi.org/10.1007/s11069-012-0234-1.

Batica, J., Gourbesville, P., 2014. Flood resilience index-methodology and application. In: International Conference on Hydroinformatics.

Blanco-Vogt, A., Schanze, J., 2014. Assessment of the physical flood susceptibility of buildings on a large scale – conceptual and methodological frameworks. Nat. Hazards Earth Syst. Sci. 14, 2105–2117. https://doi.org/10.5194/nhess-14-2105-2014.

Chen, Y.-R., Yeh, C.-H., Yu, B., 2011. Integrated application of the analytic hierarchy process and the geographic information system for flood risk assessment and flood plain management in Taiwan. Nat. Hazards 59, 1261–1276. https://doi.org/10.1007/s11069-011-9831-7.

City of Toronto, 2016a. Toronto Facts [WWW Document]. https://www1.toronto.ca/wps/portal/contentonly?vgnextoid=1d66f937de453410VgnVCM10000071d60f89RCRD&vgnextchannel=57a12cc817453410VgnVCM10000071d60f89RCRD. (Accessed 15 December 2016).

City of Toronto, 2016b. Open Data - Toronto [WWW Document]. http://www1.toronto.ca/wps/portal/contentonly?vgnextoid=9e56e03bb8d1e310VgnVCM10000071d60f89RCRD. (Accessed 20 December 2016).

City of Toronto, 2016c. Basement Flooding [WWW Document]. http://www1.toronto.ca/wps/portal/contentonly?vgnextoid=30ee7c6a9967f310VgnVCM10000071d60f89RCRD. (Accessed 19 December 2016).

City of Toronto, 2014. Wards — Toronto Ward Boundary Review [WWW Document]. http://www.drawthelines.ca/wards/. (Accessed 19 December 2016).

Cockburn, G., Tesfamariam, S., 2012. Earthquake disaster risk index for Canadian cities using Bayesian belief networks. Georisk Assess. Manag. Risk Eng. Syst. Geohazards 6, 128–140. https://doi.org/10.1080/17499518.2011.650147.

Cooper, G.F., Herskovits, E., 1992. A Bayesian method for the induction of probabilistic networks from data. Mach. Learn 9, 309–347. https://doi.org/10.1007/BF00994110.

Cutter, S.L., Burton, C.G., Emrich, C.T., 2010. Disaster resilience indicators for benchmarking baseline conditions. J. Homel. Secur. Emerg. Manag. 7 https://doi.org/10.2202/1547-7355.1732.

Environment and natural resources, 2017. Engineering Climate Datasets [WWW Document]. http://climate.weather.gc.ca/prods_servs/engineering_e.html. (Accessed 19 April 2017).

EverGreen, 2017. Greater Toronto Area Watersheds [WWW Document]. https://www.evergreen.ca/downloads/pdfs/watershed-toolkit/GTAwatersheds_FINAL.pdf. (Accessed 5 October 2017).

Foody, G.M., Ghoneim, E.M., Arnell, N.W., 2004. Predicting locations sensitive to flash flooding in an arid environment. J. Hydrol. 292, 48–58. https://doi.org/10.1016/j.jhydrol.2003.12.045.

Fratini, C., Elle, M., Jensen, M.B., Mikkelsen, P.S., 2011. A Theoretical Framework for Sustainable Transition towards Climate Change Adaptation in Urban Areas, pp. 11–16.

Fratini, C.F., Geldof, G.D., Kluck, J., Mikkelsen, P.S., 2012. Three Points Approach (3PA) for urban flood risk management: A tool to support climate change adaptation through transdisciplinarity and multifunctionality. Urban Water J. 9, 317–331. https://doi.org/10.1080/1573062X.2012.668913.

Hammond, M.J., Chen, A.S., Djordjević, S., Butler, D., Mark, O., 2015. Urban flood impact assessment: A state-of-the-art review. Urban Water J. 12, 14–29. https://doi.org/10.1080/1573062X.2013.857421.

Holling, C.S., 1973. Resilience and Stability of Ecological Systems. Annu. Rev. Ecol. Syst. 4, 1–23.

Jensen, F.V., Nielsen, T.D., 2007. Bayesian Networks and Decision Graphs, Information Science and Statistics. Springer New York, New York, NY. https://doi.org/10.1007/978-0-387-68282-2.

Joseph, S.A., Adams, B.J., McCabe, B., 2010. Methodology for Bayesian Belief Network Development to Facilitate Compliance with Water Quality Regulations. J. Infrastruct. Syst. 16, 58–65. https://doi.org/10.1061/(ASCE)1076-0342(2010)16:1(58).

Kabir, G., Tesfamariam, S., Francisque, A., Sadiq, R., 2015. Evaluating risk of water mains failure using a Bayesian belief network model. Eur. J. Oper. Res. 240, 220–234. https://doi.org/10.1016/j.ejor.2014.06.033.

Karamouz, M., Zahmatkesh, Z., Nazif, S., 2014. Quantifying Resilience to Coastal Flood Events: A Case Study of New York City. In: World Environmental and Water Resources Congress 2014. American Society of Civil Engineers, Reston, VA, pp. 911–923. https://doi.org/10.1061/9780784413548.094.

Kellershohn, D., 2016. Reducing Flood Risk in Toronto [WWW Document]. https://www.iclr.org/images/Kellerhson_ICLR_Presentation_160218_dk.pdf. (Accessed 1 October 2017).

Kim, B., Sanders, B.F., 2016. Dam-Break Flood Model Uncertainty Assessment: Case Study of Extreme Flooding with Multiple Dam Failures in Gangneung, South Korea. J. Hydraul. Eng. 142, 5016002. https://doi.org/10.1061/(ASCE)HY.1943-7900.0001097.

Klimas, R., 2009. Design Criteria for Sewers and Watermains [WWW Document]. https://www1.toronto.ca/city_of_toronto/toronto_water/articles/files/pdf/package_sewer_and_watermain_manual.pdf. (Accessed 19 December 2016).

Kron, W., 2005. Flood Risk = Hazard * Values * Vulnerability. Water Int. 30, 58–68.

Kubal, C., Haase, D., Meyer, V., Scheuer, S., 2009. Integrated urban flood risk assessment – adapting a multicriteria approach to a city. Nat. Hazards Earth Syst. Sci. 9, 1881–1895.

Lavell, A., Oppenheimer, M., Moser, S., Takeuchi, K., Oppenheimer, M., Diop, C., Hess, J., Lempert, R., Li, J., Muir, R., Stocker, F., Qin, D., Dokken, D., Ebi, K., Mach, K., Plattner, G., Allen, S., Tignor, M., Midgley, P., 2012. Climate Change: New Dimensions in Disaster Risk, Exposure, Vulnerability, and Resilience. Cambridge University Press.

Lee, M., Kang, J., Jeon, S., 2012. Application of Frequency Ratio Model and Validation for Predictive Flooded Area Susceptibility Mapping Using GIS. Ieee Igarss, pp. 895–898.

Liao, K.-H., 2012. A Theory on Urban Resilience to Floods—A Basis for Alternative Planning Practices. Ecol. Soc. 17 https://doi.org/10.5751/es-05231-170448, 48.

Maksimović, Č., Prodanović, D., Boonya-Aroonnet, S., Leitão, J.P., Djordjević, S., Allitt, R., 2009. Overland flow and pathway analysis for modelling of urban pluvial flooding. J. Hydraul. Res. 47, 512–523. https://doi.org/10.1080/00221686.2009.9522027.

Masoudian, M., Theobald, S., 2011. Influence of land surface topography on flood hydrograph. J. Am. Sci. 77, 354–361.

Nadkarni, S., Shenoy, P.P., 2001. A Bayesian network approach to making inferences in causal maps. Eur. J. Oper. Res. 128, 479–498. https://doi.org/10.1016/S0377-2217(99)00368-9.

Nasiri, H., Mohd Yusof, M.J., Mohammad Ali, T.A., 2016. An overview to flood vulnerability assessment methods. Sustain. Water Resour. Manag. 2, 331–336. https://doi.org/10.1007/s40899-016-0051-x.

Neapolitan, R.E., 2004. Learning Bayesian Networks. Pearson Prentice Hall.

Nitivattananon, V., Tu, T.T., Rattanapan, A., Asavanant, J., 2009. Vulnerability and Resilience of Urban Communities under Coastal Hazard Conditions in Southeast Asia. In: Fifth Urban Research Symposium.

Norsys Software Corp, 2016. Norsys Software Corp. - Bayes Net Software. WWW Document. https://www.norsys.com/. (Accessed 31 January 2017).

Orencio, P.M., Fujii, M., 2013. A localized disaster-resilience index to assess coastal communities based on an analytic hierarchy process (AHP). Int. J. Disaster Risk Reduct. 3, 62–75. https://doi.org/10.1016/j.ijdrr.2012.11.006.

Pearl, J., 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers.

Pradhan, B., 2010. Flood susceptible mapping and risk area delineation using logistic regression, GIS and remote sensing. J. Spat. Hydrol. 9.

Rahmati, O., Pourghasemi, H.R., Zeinivand, H., 2016. Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan Province, Iran. Geocarto Int. 31, 42–70. https://doi.org/10.1080/10106049.2015.1041559.

Saul, A.J., Djordjević, S., Maksimović, čedo, Blanksby, J., 2010. Integrated Urban Flood Modelling. In: Flood Risk Science and Management. Wiley-Blackwell, Oxford, UK, pp. 258–288. https://doi.org/10.1002/9781444324846.ch13.

Shaw, R., 2009. Climate Disaster Resilience: Focus on Coastal Urban Cities. Asian J. Environ. Disaster Manag. 1.

Stefanidis, S., Stathis, D., 2013. Assessment of flood hazard based on natural and anthropogenic factors using analytic hierarchy process (AHP). Nat. Hazards 68, 569–585. https://doi.org/10.1007/s11069-013-0639-5.

Sun, H., Cheng, X., Dai, M., 2016. Regional flood disaster resilience evaluation based on analytic network process: a case study of the Chaohu Lake Basin, Anhui Province, China. Nat. Hazards 82, 39–58. https://doi.org/10.1007/s11069-016-2178-3.

Swiss Re, 2013. The Road to Flood Resilience in Canada [WWW Document]. http://www.swissre.com/library/The_road_to_flood_resilience_in_Canada.html. (Accessed 7 February 2017).

Tang, Z., McCabe, B., 2007. Developing Complete Conditional Probability Tables from Fractional Data for Bayesian Belief Networks. J. Comput. Civ. Eng. 21, 265–276. https://doi.org/10.1061/(ASCE)0887-3801(2007)21:4(265).

Tehrany, M.S., Lee, M.-J., Pradhan, B., Jebur, M.N., Lee, S., 2014a. Flood susceptibility mapping using integrated bivariate and multivariate statistical models. Environ. Earth Sci. 72, 4001–4015. https://doi.org/10.1007/s12665-014-3289-3.

Tehrany, M.S., Pradhan, B., Jebur, M.N., 2015a. Flood susceptibility analysis and its verification using a novel ensemble support vector machine and frequency ratio method. Stoch. Environ. Res. Risk Assess. 29, 1149–1165. https://doi.org/10.1007/s00477-015-1021-9.

Tehrany, M.S., Pradhan, B., Jebur, M.N., 2014b. Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. J. Hydrol. 512, 332–343. https://doi.org/10.1016/j.jhydrol.2014.03.008.

Tehrany, M.S., Pradhan, B., Jebur, M.N., 2013. Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. J. Hydrol. 504, 69–79. https://doi.org/10.1016/j.jhydrol.2013.09.034.

Tehrany, M.S., Pradhan, B., Mansor, S., Ahmad, N., 2015b. Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. CATENA 125, 91–101. https://doi.org/10.1016/j.catena.2014.10.017.

Thomsen, N., Binning, P., McKnight, U., Tuxen, N., Bjerg, P., Troldborg, M., 2016. A Bayesian belief network approach for assessing uncertainty in conceptual site models at contaminated sites. J. Contam. Hydrol. 188.

Urban Forestry, 2009. Forest and Land Cover - Parks and Recreation - Data Catalogue | City of Toronto [WWW Document]. https://www1.toronto.ca/wps/portal/contentonly?vgnextoid=1b30790e6f21d210VgnVCM1000003dd60f89RCRD. (Accessed 20 December 2016).

Valeol, C., Rasmussen, P., 2000. Topographic Influences on Flood Frequency Analyses. Can. Water Resour. J. 25.

Van Dijk, E., van der Meulen, J., Kluck, J., Straatman, J.H.M., 2014. Comparing modelling techniques for analysing urban pluvial flooding. Water Sci. Technol. 69, 305–311. https://doi.org/10.2166/wst.2013.699.

Van Veen, B.A.D., Vatvani, D., Zijl, F., 2014. Tsunami flood modelling for Aceh & west Sumatra and its application for an early warning system. Cont. Shelf Res. 79, 46–53. https://doi.org/10.1016/j.csr.2012.08.020.

Wang, Y., Li, Z., Tang, Z., Zeng, G., 2011. A GIS-Based Spatial Multi-Criteria Approach for Flood Risk Assessment in the Dongting Lake Region, Hunan, Central China. Water Resour. Manag. 25, 3465–3484. https://doi.org/10.1007/s11269-011-9866-2.

Zou, Q., Zhou, J., Zhou, C., Song, L., Guo, J., 2013. Comprehensive flood risk assessment based on set pair analysis-variable fuzzy sets model and fuzzy AHP. Stoch. Environ. Res. Risk Assess. 27, 525–546. https://doi.org/10.1007/s00477-012-0598-5.